International Journal of Business and Society, Vol. 18 S4, 2017, 718-726

#VACCINATION: IDENTIFYING INFLUENCERS IN THE VACCINATION DISCUSSION ON TWITTER THROUGH SOCIAL NETWORK VISUALISATION

Jamali Bujang Sanawi[♣]
Universiti Malaysia Sarawak

Mus Chairil Samani
Universiti Malaysia Sarawak

Malia Taibi
Universiti Malaysia Sarawak

ABSTRACT

The emergence of social media such as Twitter has brought the discussion about vaccination to a new dimension. In this study, the researchers aim to explore the discussion on issues related to vaccination on social media platform, specifically Twitter and identify the 'influencers' in the conversation. The paper also highlighted some of the subgroups that exist within the network where the influential individuals or organisations are the centre of their attention. The findings show that there are six types of influencers that dictate the discourse on vaccination on Twitter which are: celebrity doctor, media organisations, homeopathy promoter, government and government agencies, blogger and renowned medical journal. It also found that some of the influencers have their own circle of audience while some of the influencers are sharing the same crowd.

Keywords: Vaccination; Social Media; Twitter Data.

1. INTRODUCTION

For centuries, vaccination has been acknowledged as one of the most effective preventive measures against some deadly and infectious diseases. The introduction of vaccination in medical practice has managed to save many lives and it is considered as one of the most significant achievement in modern medicine (Kata, 2012; Mercadal, n.d.; Petts and Niemeyer, 2004). Apart from preventing certain diseases, vaccination also prevent the development of antibiotic resistance to some pathogen and able to protect individual from 'bioterrorism' attack (Andre et al., 2008).

In a study conducted by Davies, Chapman and Leask (2002), they found that the probability for parents to encounter an anti-vaccination website on the Internet is much higher than pro-vaccination websites. There are numerous argumentative and rhetorical strategies that these websites often use when presenting their claims, which include denigrating scientific studies, endorsing poorly conducted studies that advocate anti-vaccination sentiments and the usage of conspiracy theories (Kata, 2012; Davies et al., 2002). The inclination of encountering vaccine-

[♣]Corresponding author: Faculty of Language and Communication, Universiti Malaysia Sarawak, 94300 UNIMAS Kota Samarahan, Sarawak, Malaysia. Email: bsjamali@unimas.my. Tel: +6082-584111.

critical websites on the Internet has raised some concern towards its effect on people's attitude towards vaccine intake. Betsch, Renkewitz, Betsch, and Ulshöfer (2010) have found that accessing vaccine-critical websites for 5 to 10 minutes do increase perception of risk of vaccinating and at the same time decreases the perception of risk of omitting vaccination. Nan and Madden (2012) also found that people who are exposed to negative blog's post on vaccination would held more negative attitude towards vaccination and reduced the intention to take vaccination.

In this study, the researchers aim to explore the discussion on issues related to vaccination on one of the most influential social media platforms, Twitter and try to identify the 'influencers' in this conversation. This paper will also highlight some of the subgroups that exist within the network where the influential individuals or organizations are at the center of attention.

2. SOCIAL MEDIA AND THE NETWORK ANALYSIS

Today, social media has become an important part of our daily lives. The platform has become ubiquitous and some of our important conversation happened on social media platform. As digital media are now increasingly connected with one another, Wellman (2001) argued that the network should not be studied in isolation because computer networks are inherently a social network that link people, organization and knowledge.

One of the most powerful and influential social media platforms is Twitter. According to Statista.com (2017), there are 328 million active Twitter users around the world as of the first quarter of 2017. Twitter also being dubbed as a type of micro blogging due to its restrictions to post only 140 characters and fewer. Murthy (2012) define micro blogging as an:

> "Internet-based service in which (1) users have a public profile in which they broadcast short public messages or updates whether they are directed to specific user(s) or not, (2) messages become publicly aggregated across users, and (3) users can decide whose messages they wish to receive, but not necessarily who can receive their messages; this is in distinction to most social networks where following each other is bi-directional (i.e. mutual)".

All these features fit the description of Twitter being a versatile, accessible and valuable tool for conveying thought and idea in a global context (McLean, Braithwaite, Swaninathan, & Stewart, 2012). Previous studies have shown that the conversations occurred on Twitter network does have its own influence on user's opinion in regards to certain topics, brands or ideologies. In a study on Twitter and branding, Jensen, Zhang, Sobel, & Abdur (2009) have found that Twitter can be an effective medium for companies and marketers to monitor customers' sentiments towards their brand and also competitors in near real time situation. The companies would receive positive brand exposure through their followers and other users who are Tweeting about their company or products. While in political arena, Tumasjan, Sprenger, Sandner, and Welpe (2010) argued that voters tweeting behavior does reflect their political preferences and almost mirrored the traditional election polls. The sentiments of the Twitters messages also coincide with the political programs, candidates' profiles and the media coverage of the campaigns.

As more people are engaging in Twitter to communicate with one another, there is an urgency to look into methodologies to study this interaction to better understand the pattern of communication

in that setting. One of the methods to study a network of connected people in online setting is through social network analysis (SNA). SNA is an analytical method that studies the structure and patterns of relationship between social actors in a network by looking at the 'nodes' and 'ties' between them (Kozinets, 2010). The nodes can be a person; organizations, ideas, messages or other concepts and the interaction between these nodes create a network. According to Hansen (2011) viewing the social world as a network can provide many insights that may not be possible to get using other approaches. SNA able to provide "detailed data about previously elusive social processes" (Smith et al., 2009) and help researchers to explore previously untouched territory in social study.

One of the key highlights in SNA is to identify the influencers in a network. In Twitter setting, influence can be defined as "the potential of an action of a user to initiate a further action by another user" (Leavitt, Burchard, Fisher, & Gilbert, 2009). This is referring to the number of retweets, mentions and replies to a particular user within a network. In a study conducted by Cha, Hamed, Benevenuto, and Gummadi (2010), they had found that mainstream news organization consistently generate high levels of retweets on various topics while celebrities on the other hand, get more mentions from their audience in comparison to other categories of users. Leavitt et al. (2009) suggested that the optimum situation for maximum influence on Twitter setting would be when the user have the most followers who have the potential to execute the most actions (retweet, mention, etc.). With many tools are made available for researchers today, such as NodeXL, Pajek, Gephi and several others, the researchers are now able to visualise a network and mapped the important individuals and organisations in the network and highlighting the subgroups within the network.

3. METHODS

This study employed a social media data mining approach using a computer program called NodeXL. It is an add-in to Microsoft Excel spreadsheet software developed by a team at the Social Media Research Foundation and it able to represent a network in a form of nodes (the programme use the term 'vertices') and edges (linkages or relationship between vertices in the network). The program is capable to automatically calculating important metrics in network analysis using formulas in graph theory. In order to determine the influencers in the network, the researchers are focusing on the top 10 nodes with the highest in-degree centrality score. In-degree centrality is used to determine the number of incoming ties (Jafar, Jalali, & Han, 2017) which in this setting, it refers to a number of mentions and replies to a particular node in this network. The higher the score means that the particular node is popular since he/she is frequently mentioned and replies to by other nodes in the network. To visualise the network, a Harel-Koren Fast Multiscale graph was generated to get the overview of the entire network and to locate the position of these influential nodes. A cluster using Clauset-Newman-Moore algorithm was also being generated to get a clearer picture on the density and the shape of the cluster in which these influential nodes belong to.

The data was extracted from Twitter on 17 July 2017 at 5:20 pm Malaysian time. The numbers of tweets are limited to the most recent 10,000 tweets to avoid an extremely massive data, which is unnecessary for the network analysis in the current study. In terms of ethical issue, Kozinets (2010) argued that "analyzing online community or culture communications or their archives is not human subject's research if the researcher can legally and easily gain access to these communications or archives". Since the researchers regarded information available on the Twitter is for public viewing, consent from the respondents involved in this analysis is deemed unnecessary. The researchers also

keep the details of the accounts as confidential thus only the description of the entities were revealed in the analysis and none of the actual username were included in the text.

4. FINDINGS

A Twitter search using '#vaccination' as a keyword in NodeXL has managed to trace 1,596 vertices or nodes in the network. Out of these nodes, there are 2,641 edges created within the network. Edges in this context are referring to the number of connection or interaction between nodes within the network of Twitter users who are searching or discussing about information related to vaccination using vaccination hashtag. These interactions can be categorized into three different relationship types, which are tweeting, replies to a tweet or mentions in a tweet. Table 1 below summarizes these relationships that were represented by these number edges in the network.

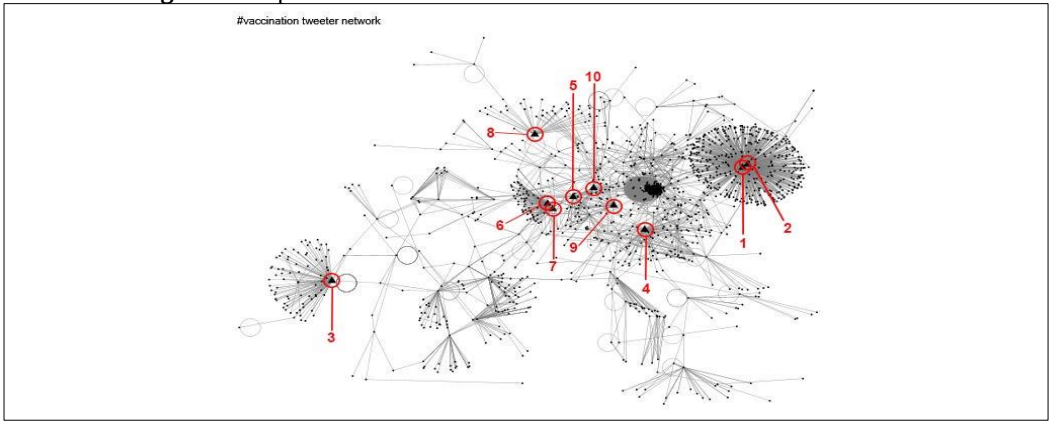
Table 1: Type of Relationship

Type of relationship	Number of edges
Tweet	421
Replies to	66
Mention	2154
Total edges	2641

From Table 1, we could see that most of the interaction (82%) involving nodes mentioning a particular node within the network in their post. Out of this, 16% involve users posting tweets with the vaccination hashtag without mentioning any names and only a small portion of the edges involving users replying to a particular tweet (2.5%).

In order to understand the structure of influence within the network, the level of in-degree centrality score was used to determine the degree of importance and influence of a particular node. Figure 1 below mapped the top 10 nodes with the highest in-degree centrality score within this network:

Figure 1: Top 10 Most Influential Vertices in the #Vaccination Network



Referring to Figure 1, the degree of influence is ranked based on the ascending order of the number with number 1 being the most influential and number 10 being the least influential within these top 10 most

important nodes in the network. The node, labelled as 1 belongs to a celebrity doctor who has his own health television program in United States. The doctor has an In-Degree Centrality score = 353, followed closely by a United States news website with In-Degree Centrality = 351 (label as number 2 in the graph). From Figure 1, both nodes seem to share the same cluster of community within the network based on their position. Both nodes also have the highest Eigenvector centrality score (both scored = .034) in comparison to the other nodes in the top 10 ranking which mean those nodes that are connected to them are also 'important' nodes because they also have many 'connections' within the network.

The third influential node (label as 3) belongs to a homeopathy promoter and organic food advocator. However, in terms of the In-Degree Centrality score, she only scores = 76 which means there is a big gap in terms of influence between her and the top two nodes in the network. The node, which is quite isolated from the media industry cluster, seems to suggest that they are dealing with quite a different crowd in this network.

The nodes that are labelled as 4, 5, 9 and 10 seemed to be a cluster from France. The nodes which labelled as 4 (In-Degree Centrality = 69) is the official Twitter account for the government of France and the one labelled as 5 is belong to France renowned newspaper organization (In-Degree Centrality = 49). Nodes labelled as 9 and 10 also share the same In-Degree Centrality score = 31, and both are from France public health agency. Their positions, which are close to each other indicate that they are sharing similar sentiment towards the issue and therefore the distance between them and the nodes that follow them are closer to each other.

Node ranked at number 6 (In-Degree Centrality = 44) belonged to a health columnist for a news organization based in Canada. This followed by a blogger who is interested in debunking pseudoscience especially related to vaccination (In-Degree Centrality = 40). A Twitter account belonged to a renowned peer-reviewed medical journal is ranked 8. Looking at the position of the node, we can say that the node also serves a particular crowd just like node 1, 2 and 3.

Using Clauset-Newman-Moore algorithm, clusters of vertices were constructed in order to have a clearer picture on the nodes and who are the crowd they interacted with. The cluster is illustrated in the Figure 2 below:

Figure 2: Top 10 Most Influential Vertices Mapped in a Cluster

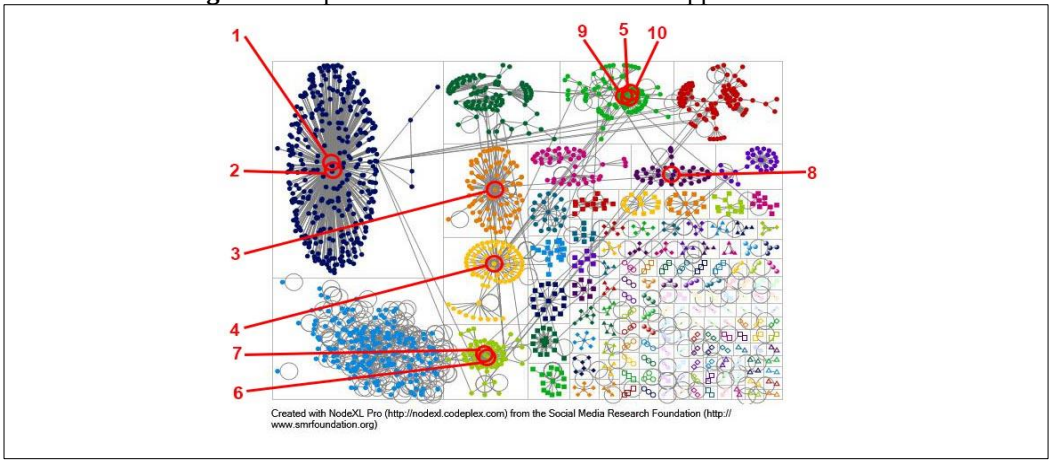


Figure 2 shows that some of those influential nodes are sharing a same cluster of crowds while some of them have crowd that unique to their own. Node 1 and 2, which are the highest scorer in terms of their In-Degree centrality in the entire network are sharing the same cluster of crowd. In comparison to other clusters, their cluster is thicker than the other which indicates that the nodes within the cluster are connected to each other and less connected to other people outside of their group (Smith et al., 2014). Both nodes are located at the heart of the cluster (also known as the hubs of the cluster), which indicate that they are the most 'influential nodes' of the cluster. Another cluster that seems to have a clear hub is the cluster where node 3 and 8 is located. This group is also well connected to each other and less connected to others in the other group. Clusters where node 4, 5, 6, 7, 9 and 10 are located however have some of the nodes isolated from the hub and being connected through nodes that act as the bridge. This means that some of these nodes require a mediator in order to get connected with the hub. Unsurprisingly the vertices from France which are node 9, 5 and 10 are clustered together except for node 4, which represent the government of France that have its own cluster. Those clusters, however, seem to have good amount of interaction based on the density of the edges that connecting some nodes in both clusters.

5. DISCUSSION

Based on the pattern in the network, it is clear that the media and individual with celebrity status are playing a major role in shaping people's opinion and idea about certain issues including health information. One of the reasons that could explain why the media and celebrity doctor (node 1 and 2 respectively) is able to gain so much attention from the nodes in the analyzed network is because of their popularity of their existence in a social network media. They quickly referred to as a reliable and immediate source of information on vaccination. They played the same role in the physical world as well as in the cyber social media realm such as Twitter. Those nodes that attached themselves with these two entities are probably the fans of the doctor and the subscribers of the news websites who happened to be interested in the issues that evolve around vaccination at that point of time.

In contrast, the homeopathy promoter seems to be offering a different view on vaccination. This is indicated based on their position and distance from each other. The homeopathy promoter is critical towards vaccination and belief in natural healing methodology, which was reflected in some of her tweets and replies in the network. Her view managed to attract a cluster of people who are inclined towards anti-vaccination within this network. Based on her In-degree Centrality score and the shape of the cluster, it seems to suggest that she is a 'prominent figure' in the cluster where a lot of mentions and replies were directed to her, instead of her toward others. A clear central node in the cluster also indicates that the people in the group relying on a particular source (the homeopathy promoter) for information thus suggesting that the group is actually consists of people who have a similar sentiment towards vaccination.

Interestingly, the two dominant clusters in the network are consists of Twitter users who reside in United States. This reflects the dynamic of opinions towards vaccination in United State whereby there are groups who are supportive towards vaccination and also groups who oppose vaccination as a practice for immunization. This pattern also shows that Americans are actively engage with social media such as Twitter in getting and circulating information related to health issues. Being a country with the largest active Twitter users in the world (Statista.com, 2017), it is unsurprising

to see how this platform has become an integral part of their life including as a source for health information.

Another interesting finding in this study is the emergence of the clusters from France who are pro-vaccination. There were two prominent clusters from France in the research network. One cluster is relying on the government while the other is relying on both the public health agencies and French media for information relating to vaccination. The emergence of clusters from France perhaps driven by the measles outbreak in Europe that had killed 35 people as the disease spreads among unvaccinated children (Forster, 2017). Similar to the US, the French are also actively using social media to get information on issues related to health and in this case, the French Twitter users are relying on two different sources with some of them relying on the government while the other relying on the public health agencies and French media for the latest development on diseases caused by unvaccinated individuals.

A blogger and a columnist also share quite a big amount of importance in this network. As shown in Figure 2, both nodes seem to appeal to a certain clusters of people in the network. This particular cluster probably interested with issues related to vaccination but prefer it to be written in a particular way. The blogger who like to be regarded as pseudoscience debunker and who is also critical towards the anti-vaccination movement often posed questions in his Twitter posts and provide a link to the whole article that he wrote on that particular topic. Instead of sharing articles from other sources which is a common practice in the other clusters, he prefers to lead the readers to his blog and read the article that he wrote using his own perspective and argumentative evidences. It would be acceptable to regard this cluster as a group of blogger and blogger followers because it seems that the way the information is presented to them is the determining factor on why they are in the cluster.

Last but not least is the cluster that is formed with a medical journal as the hub that connecting all the other nodes in the cluster. As shown in Figure 2, the cluster is not as dense as the other clusters that previously discussed and the location of the cluster also a little bit isolated from the rest. This indicates that the cluster's members are quite exclusive and probably consist of medical scholars, practitioners and students who use this journal as the main reference for updates on medical developments including those related with vaccination. Since the cluster seems to be quite exclusive to a certain group of people, with the journal as the central hub of the cluster does imply a different kind of conversation that this group must have. Possibly most of the issues discussed in this cluster are in the form of an intellectual and academic discussion in which each member has their own interest in it. For the journal to emerge as one of the most influential node in this network at this particular time frame shows that the current issue that evolve around vaccination does not only draw interest among the regular public but also scholars and researchers in the field of medicine as well.

6. CONCLUSION

In conclusion, this study has shown some of the characteristics of influencers in the conversation about vaccination on Twitter. From the findings and discussion, we can conclude that there are six types of influencers that dictate the discourse on vaccination on Twitter. They are celebrity doctors, media organisations, homeopathy promoters, government departments and agencies, bloggers and renowned medical journals. For each of the influencers they have a different type of clusters that

attracted to them in which each of the clusters have a unique feature themselves. The illustration of the network in a visual form allows us to see who and where the influencers are located within the network. It also projects the size of the group that they have influenced on. This finding supports the model on blog-mediated crisis communication that was developed by Liu, Jinb, Brionesa, and Kucha. (2012).

This study however is not without its own limitations. The dynamic and fluid nature of the Twitter environment will make it difficult to generalize the findings in the long term. A re-analysis will have to be done from time to time to see the changes that occur with the influencers and the nodes created within the social network. Thus, the visualise network is situated within the spatio-temporal structure at the time the data was extracted from Twitter website. Replicating this similar study in the future may yield a different result due to the different space and time continuum. It should be noted that this is quantitative study that is done by interpreting the mathematical calculation derived from NodeXL. The researchers argue that the network analysis alone will not be sufficient to explain about the people and the organization within a social network. The analysis should be triangulated with a study using qualitative approach such as thematic or semantic analysis so as to anticipate the existence of gaps in social network analysis.

REFERENCES

Andre, F. E., Booy, R., Bock, H. L., Clemens, J., Datta, S. K., John, T. J., Lee, B. W., Lolekha, S., Peltola, H., Ruff, T. A., Santosham, M., & Schmitt, H. J. (2008). Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bulletin of the World Health Organization*, *86*(2), 140-146.

Betsch, C., Renkewitz, F., Betsch, T., & Ulshöfer, C. (2010). The influence of vaccine-critical websites on perceiving vaccination risks. *Journal of Health Psychology, 15*(3), 446–455.

Cha, M., Hamed, H., Benevenuto, F., & Gummadi, K. P. (2010). *Measuring user influence in Twitter: The million follower fallacy.* Paper presented at the Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 23-26 May (pp. 10-17). Washington, DC, USA: ICWSM.

Davies, P., Chapman, S., & Leask, J. (2002). Antivaccination activists on the world wide web. *Archives of Disease in Childhood*, *87*, 22-25.

Forster, K. (2017, July 12). Measles kills 35 people in Europe as disease spreads through un-vaccinated children, warns WHO. *The Independent*. Retrieved from http://www.independent.co.uk/news/health/measles-deaths-europe-outbreak-vaccination-mmr-who-unacceptably-tragedy-italy-romania-mmr-anti-a7836581.html

Hansen, D. L. (2011). Exploring social media relationships. *On the Horizon*, *19*(1), 43-51.

Jafar Jalali, S. M., & Han, W. P. (2017). Conversations about open data on twitter. *International Journal of Contents*, *13*(1), 31-6.

Jensen, B. J., Zhang, M., Sobel, K., & Abdur, C. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of The American Society for Information Science and Technology*, *60*(11), 2169–2188.

Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm – An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, *30*, 3778– 3789.

Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*. London: SAGE.

Leavitt, A., Burchard, E., Fisher, D., & Gilbert, S. (2009). The Influentials: New approaches for analyzing influence on Twitter. *Web Ecology Project*. Retrieved from http://tinyurl.com/lzjlzq

Liu, B. F., Jinb, Y., Brionesa, R., & Kucha, B. (2012). Managing turbulence in the blogosphere: Evaluating the blog-mediated crisis communication model with the American red cross. *Journal of Public Relations Research, 24*(4), 353-370

Maclean, F., Jones, D., Carin-Levy, G., & Hunter, H. (2013). Understanding Twitter. *British Journal of Occupational Therapy*, *76*(6), 295-298.

McLean, K. L., Braithwaite, M., Swaninathan, J., & Stewart, A. (2012). Variability in control of onion white rot by *Trichoderma atroviride* under different disease pressures. *Australasian Plant Pathology, 41*(4), 341-346.

Mercadal, T. (n.d.). Anti-vaccination movement. *Research Starters Sociology*, 2-7.

Murthy, D. (2012). Towards a sociological understanding of social media: Theorizing Twitter. *Sociology*, 1-15.

Nan, X., & Madden, K. (2012). HPV vaccine information in the blogosphere: how positive and negative blogs influence vaccine-related risk perceptions, attitudes, and behavioral intentions. *Health Communication*, *27*(8), 829-836.

Petts, J., & Niemeyer, S. (2004). Health risk communication and amplification: learning from the MMR vaccination controversy. *Health, Risk and Society*, *6*(1), 7-23.

Smith, M. A., Rainie, L., Shneiderman, B., & Himelboim, I. (2014, February 20). *Mapping Twitter topic networks: From polarized crowds to community clusters*. Retrieved from http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Rodrigues, E. M., Barash, V, Dunne, C., Capone, T., Perer, A., & Gleave, E. (2009). *Analyzing (social media) networks with NodeXL*. Retrieved from https://www.cs.umd.edu/~ben/papers/Smith2009Analyzing.pdf

Statista.com. (2017). *Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2017 (in millions)*. Retrieved from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). *Predicting elections with Twitter: What 140 characters' reveal about political sentiment.* Paper presented at the Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, USA, 23-26 May (pp. 175-185). Washington, DC, USA: ICWSM.

Wellman, B. (2001). Computer networks as social networks. *Science*, *293*, 2031-2034.